

TARTU ÜLIKOOL

MATEMAATIKA-INFORMAATIKATEADUSKOND

Arvutiteaduse instituut

Infotehnoloogia õppekava

Lauri Tammeveski

**Eestikeelse jalgpalli-alase tekstikorpuse automaatne märgendamine**

**jalgpalli-alase Frameneti abil**

Bakalaureusetöö (6 EAP)

Juhendaja: teadur Neeme Kahusk

Tartu 2014

Eestikeelse jalgpalli-alase tekstikorpuse automaatne märgendamine jalgpalli-alase Frameneti abil

Lühikokkuvõte:

Käesoleva töö eesmärgiks on uurida ning üritada lahendada eestikeelse teksti automaatse freimidega märgendamise probleemi. Üldine eestikeelne Framenet on alles algusjärgus, kuid olemas on terviklik jalgpalli-alane freimide ressurss, mille abil üritame tõestada hüpoteesi, et jalgpalli-alase teksti märgendamiseks piisab vaid morfoloogilisest ning süntaktilisest infost. Sellele hüpoteesile me siiski kinnitust ei saanud, kuna sama tähendust kandvat lauset on võimalik esitada liiga paljudel erinevatel viisidel. Lisaks täiendasime jalgpalli-alaste sõnadega Eesti suurimat leksikaal-semantilist andmebaasi, Wordnetti.

Võtmesõnad: freim, automaatne märgendamine, Framenet, Wordnet

Estonian football specific corpora automatic semantic role labeling with football specific Framenet

Abstract:

Research and a possible solution to the problem of automatic semantic role labeling of text in Estonian is carried out in this paper. A general Estonian Framenet is in the starting phase, but there is also available a football specific Framenet. We try to prove the hypothesis that morphological and syntactical information is enough for automatic semantic role labeling in football related corpora. Unfortunately, we did not achieve a confirmation for the hypothesis, because there are too many ways to present sentences that have the same meaning. In addition, we supplemented Estonian biggest lexical-syntactic database with football related words.

Keywords: frame, automatic semantic role labeling, Framenet, Wordnet

Autorideklaratsioon

Deklareerin, et käesolev bakalaureusetöö on minu iseseisva töö tulemus, on esitatud Tartu Ülikooli Arvutiteaduse instituudi lõpudiplomi taotlemiseks infotehnoloogia erialal. Lõputöö alusel ei ole varem eriala lõpudiplomit taotletud.

Autor Lauri Tammeveski, 14.05.14

# Sisukord

Sissejuhatus.....	4
1. Taustinfo.....	6
1.1. Freimid.....	6
1.2. FrameNet.....	7
1.3. Kasutatav eestikeelne jalgpalli-alane Framenet.....	8
1.4. Kasutatav eestikeelne jalgpalli-alane korpus.....	8
1.5. Teksti märgendamine freimidega.....	10
1.5.1. Käsitsi märgendamine.....	10
1.5.2. Automaatne märgendamine.....	11
1.5.3. Automaatse märgendamise võimalikud probleemid.....	13
1.6. Wordnet.....	14
2. Praktilise töö ülevaade.....	15
2.1. Frameneti täiendamine.....	15
2.2. Automaatne märgendamine.....	17
2.3. Eesti Wordneti täiendamine.....	22
Kokkuvõte.....	23
Kasutatud kirjandus.....	24
Lisad.....	27
Lisa 1.....	27
Lisa 2.....	28
Lisa 3.....	39

## Sissejuhatus

Kõnekeele ning tekstide arusaadavaks tegemine arvutitele on vähemalt pool sajandit olnud teadlaste suur eesmärk. Kui morfoloogiliselt (sõnaliikide määramine) ning süntaktiliselt (sõnade järjekorra ja lausete struktuuri määramine) on tekstide analüüs juba üsna täpne, siis semantiline ehk tähenduslik analüüs on veel problemaatiline.[1]

Üks võimalus selleks on kasutada freimiteooriat, mida selles töös tehaksegi. Kuna sõnast üksinda ei piisa tema tähenduse mõistmiseks, siis vaadatakse tervet lauset kui situatsiooni ning sellest luuakse freim, mille puhul on tegemist andmestruktuuriga, mis kirjeldab seda lauset.[2]

Õige freimi ja selle elementide lisamist nimetatakse selle teksti märgendamiseks. Märgendamine on võimalik nii käsitsi kui ka automaatselt.[1]

Kõige rohkem on freimidega tegeletud Berkeley Rahvusvahelises Arvutiteaduste Instituudis, kus on loodud FrameNeti nimeline projekt. Selles projekti raames on loodud üle 800 freimi ning 150 000 näitelause kasutades käsitsi märgendamist. Peaaegu kõik automaatse märgendamisega seotud teadustööd põhinevad FrameNeti näitelause alusel tehtava masinõppega, mis annavad umbes 65% täpsuse.[3, 4]

Ka eesti keeles on Framenetiga algust tehtud, kuid see on alles algusjärgus. Samas on eesti keeles loodud terviklik jalgpalli-alane freimide võrgustik ja tekstikorpus. Käesoleva töö eesmärk ongi üritada automaatselt selle korpuse tekste märgendada.[5, 6]

Antud korpus ehk tekstide kogum on morfoloogiliselt ning süntaktiliselt analüüsitud ning selle tööga üritame kinnitust saada hüpoteesile, et jalgpalli-alaste tekstide puhul piisab nende morfoloogilisest ning süntaktilisest infost ning freimide täiustamisest sarnase infoga, et märgendada tekste masinõppele sarnase täpsusega. Eduka automaatse märgendamise puhul oleks võimalik luua süsteeme, mis oleksid edukad sisukokkuvõtete tegemisel, küsimustele vastamisel, infootsingul, tekstide omavahelisel võrdlemisel ning masintõlkel.

Lisaülesandeks on Eesti Wordneti täiustamine jalgpalli-alaste väljenditega, mis leiduvad Framenetis, kuid mida ei ole veel Wordnetis, või mis on seal, aga ei oma vastavat tähendust. Tegemist on Eesti kõige arenenuma leksikaalse andmebaasiga ning seetõttu tasub seda alati

täiendada sõnadega, mida seal veel ei ole. Sellest võivad hiljem kasu saada nii tavakasutajad kui ka inimesed, kes kasutavad seda ressursi tõlkesüsteemide või muu arendamiseks.

Töö on jaotatud kaheks peatükiks. Esimese osa moodustab teoreetiline ülevaade freimidest ja nende märgendamisest ning samuti kirjeldatakse kasutatavaid ressursse. Teises peatükis on kirjas, mida tuli teha enne automaatse märgendamise juurde asumist, millisel viisil toimus märgendamisprogrammi loomine ning kuidas leiti sõnu, mis täiendaks Eesti Wordnetti.

Töö praktiline tulemus on ligipääsetav lehel [www.github.com/Tamme/jalgpalliFreimNet](https://www.github.com/Tamme/jalgpalliFreimNet). Lisades asuvad loend loodud failidest, käsitsi märgendatud tekst ning Wordneti täiendamiseks sobivad sõnad.

# 1. Taustinfo

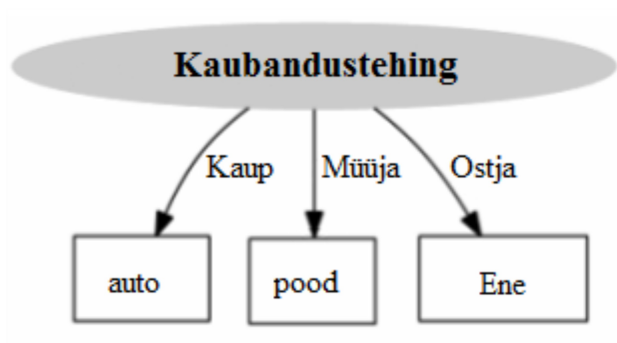
## 1.1. Freimid

Freime tutvustati esmakordselt 1975. aastal ning aasta hiljem avaldati nendest pikem artikkel. Kuna sõna tähendusest ei ole võimalik aru saada ilma selle sõna kohta käiva infota, siis loodi freim – mingit sündmust või situatsiooni kirjeldav andmestruktuur, kus on kirjas sõnad või mõisted, mis vastava freimi välja kutsuvad (nimetatakse leksikaalseteks üksusteks, edaspidi LÜ), elemendid, mis sellel freimil on, ning seosed teiste freimidega. See toimib vahelülina semantika ning teksti vahel.[1]

Võtame kolm näitelause:

- Ene ostis poest auto.
- Pood müüs Enele auto.
- Ene tegi poest auto ostu.

Need kõik kirjeldavad sama situatsiooni ning tänu sellele on ka sama freimi "Kaubatehing" esilekutsujad, mis on kujutatud joonisel 1. Sellelt on näha lisaks freimi nimele ka selle elemendid: müüja – pood, ostja – Ene ning kaup – auto. Kujutatakse seda tavaliselt järgnevalt: [ostja Ene] **ostis** [müüja poest] [kaup auto] [9].



Joonis 1. Freim "Kaubatehing".

Tänu sellisele üldistamise võimalusele tulevad välja ka freimisüsteemide peamised kasutusala – kokkuvõtete tegemine, tekstide omavaheline võrdlemine, otsingud ning masintõlge [1].

## 1.2. FrameNet

FrameNet on Berkeley Rahvusvahelises Arvutiteaduste Instituudis algatatud projekt, mille eesmärgiks on luua võrgus kättesaadav ingliskeelne leksikaalne ressurss, mis põhineb freimiteoorial. Seal üritatakse dokumenteerida iga sõna kõiki erinevaid tähendusi ning süntaktilisi esitusviise.[4]

Töö käib viisil, kus analüüsitakse järjest igat freimi, antakse sellele definitsioon, vaadatakse millised LÜ-d võivad selle välja kutsuda ning otsitakse korpusest (põhiliselt 100 miljoni sõnaline Briti Riiklik korpus (ingl k British National Corpus)) näitelauseid iga selle LÜ tähenduse või süntaktilise esinemisviisi kohta vastava freimi suhtes. Seejärel märgendatakse see lause käsitsi ning lisatakse andmebaasi. Eraldi osa moodustab terviktekstide märgendamine, mis on eriline, kuna ühes lauses võib olla erinevaid freime välja kutsuvad LÜ-d.[4]

Hetkel on loodud andmebaasis üle 10 000 leksikaalse üksuse, 800 hierarhilise freimi ning 150 000 märgendatud näitelause. Sellest tulenevalt on FrameNet suurim ja enim kasutatud ingliskeelne freimidefinitsioonide ressurss.[4, 10]

Kuna igal freimil on oma semantiline tähendus, siis kokkuvõttes grupeeritakse sarnase tähendusega sõnad kokku, mis aitavad hiljem semantilisel analüüsil.[4]

Kuigi freimi elemendid on erinevates keeltes üsna sarnased, võib siiski olla ka erinevusi ning seetõttu on loodud ka näiteks jaapani, hispaania ja saksa keele FrameNeti versioone [1].

Ka eestikeelse freimileksikoniga on algust tehtud – kuuesaja ingliskeelse freimi leksikaalsed üksused on tõlgitud eesti keelde [6]. Märgendatud näitelauseid eriti veel ei ole, leiduvad vaid mõned liikumist väljendavad laused [7]. Lisaks on olemas verbide freimileksikon, kus on ka näitelauseid, kuid see ei ole piisavalt mahukas, et terviktekste märgendada [8]. Kolmanda Framenetina on eesti keeles jalgpalli-alaste mõistete ning situatsioonide oma, milles ei ole aga märgendatud näitelauseid, vaid ainult freimid [5]. Seda me antud töös kasutama hakkamegi, kuna selle abil võiks olla võimalik ka antud valdkonna terviktekste märgendada. Edaspidi kasutame sõna Framenet, viidates mõnele üldisele freimileksikonile, ning nime FrameNet, kui mõtleme ingliskeelset Berkeley ressurssi.

### 1.3. Kasutatav eestikeelne jalgpalli-alane Framenet

Eesti keeles on freimidega vähe tegeletud ning teadaolevalt on ainult üks suurem terviklik Framenet. See on valdkonnaspetsiifiline ning kirjeldab kõiki jalgpalli-alased mõisteid ning termineid. Tegemist on veebilehel [www.kicktionary.de](http://www.kicktionary.de) asuvast ingliskeelsest Framenetist, mis samuti kirjeldab jalgpalli-alaseid väljendeid, tõlgitud ressursiga. Selles eestikeelses Framenetis on 107 freimi, mis on loodetavasti juba piisavalt suur hulk, et üritada ühe kindla valdkonna korpust nendega märgendada.[5]

Framenet asub XML failis ning on joonisel 2 kujutatud struktuuriga, kus "framelexicon" on kõiki freime siduv ülemmärgend, "frame" on ühte freimi tähistav märgend, "LexicalUnits" seob kõik freimi välja kutsuvad leksikaalsed üksused, "lu" tähistab ühte LÜ-d, "Elements" ühendab kõiki freimi elemente ning "element" näitab ühte freimi elementi. Lisaks on freimidel atribuudid "name", mis tähistab freimi või elemendi nime, "lemma", mis on LÜ nimi ning "optional", mis näitab, kas element on lauses kohustuslik või mitte. Leksikaalseid üksusi võib olla 1 või rohkem ning elemente 0 või rohkem. Tegemist on standardse XML formaadiga.[5, 11]

```
<framelexicon>
  <frame name="freimi nimi">
    <LexicalUnits>
      <lu lemma="lemma nimi" />
      ...
    </LexicalUnits>
    <Elements>
      <element name="elemendi nimi" optional="true" />
      ...
    </Elements>
  </frame>
</framelexicon>
```

Joonis 2. Frameneti struktuur.

### 1.4. Kasutatav eestikeelne jalgpalli-alane korpus

Hüpoteeside kontrollimiseks on lisaks Framenetile vaja ka korpust, mille tekste saaks analüüsida. Koos jalgpalli-alase Framenetiga on ka vastav korpus loodud, mis on morfoloogiliselt ning süntaktiliselt analüüsitud, kasutades programmi EstCGParser [5].



Joonisel 3 on kujutatud üks morfoloogiliselt ning süntaktiliselt analüüsitud lause. Igas reas on ühe sõna analüüsi tulemus, kus esmalt on tekstis esinenud sõna, seejärel tema algvorm, millele on pluss märgi abil lisatud käände- või pöördelõpp. Järgmisena on kahe kahekordse kaldkriipsu vahel morfoloogiline info ning viimasena süntaktiline info. Kui sõna kuju järgi ei ole võimalik aru saada, mis algvormiga on tegu (näiteks kas "lõi" tähistab "looma" või "lööma"), siis tuleb pärast süntaktilist infot teise algvormi võimalus ja selle morfoloogiline ning süntaktiline info ning nii edasi, kuni kõik võimalikud algvormid on kirjas. Seda on täpsemalt kirjeldatud allikas [12].

Kasutatavad märgendid on järgmised:

Süntaktilised märgendid, mis on põhiliselt kasutuses [13]:

- @SUBJ – alus ehk subjekt
- @OBJ – sihitis ehk objekt
- @ADVL – määrus ehk adverbiaal
- FMV – öeldis
- FCV – olema vorm

Morfoloogilised märgendid, mis on põhiliselt kasutuses [14, 15]:

- Sõnaliikide jaoks:
  - \_A\_ - omadussõna, nt kallis
  - \_D\_ - määrsõna (adverb), nt kõrvuti
  - \_K\_ - kaassõna (pre / postpositsioon), nt kaudu
  - \_J\_ - sidesõna nt ja
  - \_N\_ - arvsõna (numeraal), täpsemalt põhi- ja järgarvsõnad (\_N\_ card ja \_N\_ ord)
  - \_S\_ - nimisõna (substantiiv), täpsemalt üld- ja pärisnimed (\_S\_ com ja \_S\_ prop)
  - \_V\_ - tegusõna (verb), täpsemalt põhi-, modaal- ja abiverbid (\_V\_ main, \_V\_ mod, \_V\_ aux), nt lugema
- Käänete jaoks: nom – nimetav, gen – omastav, part – osastav, ill – sisseütlev, in – seesütlev, el – seestütlev, all – alaleütlev, ad – alalütlev, abl – alaltütlev, tr – saav, ter – rajav, es – olev, ab – ilmaütlev, kom – kaasaütlev

- Arvude jaoks: sg – ainsus, pl – mitmus

Rida algusega \$LA\$ tähistab lause algust ning \$LL\$ lause lõppu. Morfoloogilisi ja süntaktilisi märgendeid on veel, kuid neid selles töös analüüsiks ei kasutata.

```
$LA$      ### @???
Eelmisel   eelmine+1 //_A_pos sg ad #cap // **CLB @AN>
nädalal    nädal+1 //_S_com sg ad // @ADVL
lõi        loo+i //_V_main indic impf sg // @+FMV   lõõ+i //_V_main indic impf sg
Taani      Taani+0 //_S_prop sg gen #cap // @OBJ @NN>
esiliigas  esi_liiga+s //_S_com sg in // @ADVL
tähtsa     tähtis+0 //_A_pos sg gen // @AN>
värava     värav+0 //_S_com sg gen // @OBJ @NN>
Kaimar     Kaimar+0 //_S_prop sg nom #cap #? // @SUBJ @NN>
Saag       saag+0 //_S_com sg nom #cap // @SUBJ
$.         . //_Z_Fst //
$LL$      ### @???
```

Joonis 3. Analüüsitud näidislause.

## 1.5. Teksti märgendamine freimidega

Lause tähenduse määramine on keeruline ülesanne. Arvestada tuleb erinevate süntaktiliste lausekonstruktsioonidega, mis võivad tegelikult edasi kanda sama mõtet. Üheks võimaluseks on lause komponentidele semantiliste rollide määramine [16]. Seda eesmärki täidabki teksti märgendamine freimidega (ingl k *frame labeling* või *frame tagging* või *sentence annotating* või *semantic role labeling* või *semantic parsing*).

Märgendamist on tehtud nii käsitsi kui ka automaatselt, enim ingliskeelsete tekstide jaoks [4, 16]. Siiani on enim edu toonud meetod, kus kõigepealt tehakse seda piisavalt suure korpuse jaoks käsitsi ning seejärel üritatakse nende kui näidete põhjal kõiki teisi tekste automaatselt märgendada [4].

### 1.5.1. Käsitsi märgendamine

Kõige rohkem on käsitsi märgendamist tehtud ja tehakse eeldatavalt ka edaspidi Berkeley FrameNetis. Seal märgendatakse laused freimide ja nende elementidega leksikograafide poolt. Lisaks kasutatakse järjepidevuse halduse süsteemi, mis peaks tagama, et ei oleks vastuolusid ega vigu. Käsitsi märgendamine on vajalik suure korpuse saamiseks, kus oleks kõik laused kindla reeglistiku alusel märgendatud. Tulemuseks on erinevad andmekihid teksti

peal. Eraldi kihid moodustavad freimi või LÜ nimed, freimi elemendid, grammatilised funktsioonid ning fraasi tüübid (nimisõna fraas, verbi fraas jne).[4]

Märgendamise reeglid, mida kasutatakse [4]:

- Märgendada tohib ainult otseselt LÜ-st või elemendist sõltuvaid sõnu. Kontekstile viitamist ei tohi arvestada.
- Märgendama peab tervet fraasi, mitte ainult selle põhiosi. Näiteks lause "mees jooksis piki väljakut" puhul on asukohaks piki väljakut, mitte lihtsalt väljak.
- Iga lause märgendatakse LÜ-st sõltuvalt.
- Tervikteksti märgendamise puhul tuleb märgendada iga LÜ ja element, isegi kui need ei anna mingit leksikograafilist või süntaktilist infot.
- Isegi kui ühe LÜ jaoks lause märgendamisel leidub sõnu, mis on LÜ-ks mõnele teisele freimile, jäetakse need märgendamata.
- Jpt.

Saadud korpust saab seejärel kasutada automaatseks märgendamiseks masinõppel põhinevatel meetoditel. Seda juba tehaksegi paljude teadlaste ning üliõpilaste poolt.[4]

### 1.5.2. Automaatne märgendamine

Automaatse märgendamise korrektsuse hindamiseks on kaks suurus: täpsus (ingl k *precision*) – kui palju märgendatud sõnadest on õigesti märgendatud ning saagis (ingl k *recall*) – kui palju sõnadest on ära märgendatud [17]. Need on suhteliselt vastandlikud, kuna tavaliselt suure täpsuse korral märgendatakse vähe sõnu või vastupidi. Semantilise analüüsi tegijatel on eesmärk need mõlemad saada võimalikult kõrgeks.

Automaatsel märgendamisel on kasutatud mitmeid erinevaid lahendusi.

Vananenud viisid on ühtlustamisel põhinevad (ingl k *unification-based*) grammatikad, nagu näiteks 1994. aastal arendatud HPSG (põhjakeskne fraasistruktuurigrammatika, ingl k *head-driven phrase structure grammar*). Selline viis põhineb käsitsi loodud grammatikal, mis nõuab sõna iga süntaktilise kasutusviisi ette nägemist. Selline meetod on aeganõudev ja tõenäoliselt piiratud kasutusalaga.[3]

Uuemad ning edukamad on masinõppel põhinevad rakendused, kus näitelause järgi õpitakse kõik reeglid ning nende põhjal märgendatakse uusi lauseid. Õppimiseks kasutatakse põhiliselt Berkeley FrameNeti projekti näitelauseid, kus neid on 150 000.[10]

Aastal 2002 avaldatud Gildea ja Jurafsky artikkel on teedrajav statistilise masinõppe valdkonnas [18, 19]. Selle puhul üritatakse iga lause jaoks näitelausest kogutud statistika põhjal määrata kõige sobivam freim [3]. Kasutusel on erinevaid mudeleid. Näiteks maksimaalse entroopia oma puhul valitakse freim, mis sobib kõikide eeldustega, aga on muidu võimalikult üldine [20]. Statistilise analüüsi täpsuseks tuli Gildea ja Jurafsky töös 65% ning saagiseks 61% [3].

Praeguse hetke parim märgendaja on SEMAFOR, mis töötati välja aastal 2010 ning see kasutab tõenäosuslikku masinõpet, mis on statistilise masinõppe edasiarendus [21]. SEMAFOR-i tööpõhimõte on järgmine [18, 19]:

- Tekst eeltöödeldakse (lisatakse sõnaliikide märgendid programmiga MXPOST, kasutatakse sõltuvusgrammatika parserit MST, ingliskeelse WordNeti abil määratakse tekstis sõnade algvormid ning iga verbi puhul märgitakse, kas tegemist on aktiivse või passiivse kõnega).
- Edasine analüüs jagatakse kolmeks: freimi väljakutsuva sõna tuvastamine, õige freimi määramine (välja kutsuv sõna võib olla mitmetähenduslik või vastavat freimi ei olegi) ning freimielementide tuvastamine.
- Freimi väljakutsuva sõna tuvastamine toimub kindla reeglistiku alusel millega saavutatakse 89,92% täpsus ning 70,79% saagis.
- Õige freimi identifitseerimisel kasutatakse tõenäosuslikku mudelit, kuna siis suudetakse analüüsida ka sõnu, mis ei ole otseselt Frameneti LÜ-de nimekirjas. Mudeli abil saavutatakse 69,75% täpsus ning 54,91% saagis.
- Freimielementide tuvastamiseks kasutatakse mudelit, mis arvestab süntaktilise analüüsi tulemust, pöörd sõnade tegumoodi, sõnade järjekorda ning kaugust LÜ-st ning sõnaliike. Sellise meetodiga saavutatakse keskmiselt 80% täpsus ning 67% saagis.

- Kokkuvõttes saadi varasemaga võrreldes efektiivsem rakendus, mis on umbes viis protsenti parema täpsuse ning saagisega.
- SEMAFOR on kätte saadav aadressilt <http://www.ark.cs.cmu.edu/SEMAFOR>.

Pärast SEMAFOR-i avalikustamist on see võetud kõigi järgmiste märgendajate võrdluse aluseks. Tihti üritatakse seda lihtsalt kiiremaks ning täpsemaks arendada. Rohkem infot edasiarenduste kohta on allikas [21].

Kuna eesti keeles on freimidega veel vähe tegeletud ja puudub suur märgendatud näitelauseite korpus, siis ei saa masinõpet kasutada. Seetõttu üritataksegi selle töö raames morfoloogilise ja süntaktilise info (edaspidi MSI) abil jalgpalli-alaseid tekste automaatselt märgendada. See meetod sarnaneb pisut käsitsi loodud grammatikale ning teistele viisidele, kus luuakse mingi reeglistik, mille järgi lauseid analüüsitakse. Täpselt sellist protseduuri uuritud kirjanduse järgi eesti keeles mõnele Framenetile tuginedes aga ei ole katsetatud ning kuna erinevate keelte analüüsil esinevad erinevad probleemid, siis on sellise hüpoteesi kontrollimine õige samm selleks, et teada saada, mis viisil eestikeelseid tekste freimidega peaks märgendama.

### **1.5.3. Automaatse märgendamise võimalikud probleemid**

Automaatne märgendamine on vägagi keeruline ülesanne. Suurimaks probleemiks on see, et FrameNet on poolik ning sealt on puudu freime ning käsitsi märgendatud näitelauseid [18]. Seetõttu leidub tekstides, mida märgendada soovitakse, sõnu ning väljendeid, millel puudub õige freim või lihtsalt ei suudeta vastavat freimi leida, kuna andmebaasis puudub vastava lausekonstruktsiooniga näitelauseid.

Lisaks on puuduseks see, et paljudel sõnadel on mitmeid erinevaid tähendusi. Seega on raske aru saada, milline tähendus kindlas lauses tuleb määrata. Olemas on ka väga üldisi elemente nagu aeg ja koht, millele vastavusse sobib väga palju sõnu ning fraase. Võimalus on ka, et lauses puudub mõni lauseosa, näiteks alus, mis teeb märgendamise keerulisemaks.[16]

Kuna iga uuritav sõna peab esinema ka näitelauseites, siis peaks näitelauseid edaspidi täiendama ning parandama, kui lisandub uusi sõnu või tähendusi [3].

Lisaks sellele ei ole teada, et keegi oleks märgendamisel arvestanud võimalusega, et tuleks luua uus freim mingi situatsiooni kohta. Siiani on üritatud kõige sarnasema freimiga märgendada.[18]

## 1.6. Wordnet

Wordnet on leksikaal-semantiline andmebaas, kus sõnad on grupeeritud sünonüümi- ehk sünohulkadeks. Iga hulk koosneb ühte ja sama mõistet väljendavatest sõnadest. Sünohulgad on omavahel ühendatud viidetega, mis vastavad semantilistele või leksikaalsetele suhetele. Enim kasutatud suheteks on hüperonüümia ning hüponüümia (vastavalt kas üks sõna on teise ülemmõiste või alammõiste). Lisaks on kasutusel holonüümia ja meronüümia (vastavalt kas sõna väljendab tervikut osade suhtes või vastupidi), antonüümia (vastandsõnad) jt.[22, 23]

Inglisekeelse WordNeti loomine algas aastal 1985, mis on FrameNetist 12 aastat varem. Sellest tulenevalt on seal juba umbes 150 000 sõna võrreldes FrameNeti 10 000 LÜ-ga [24].

Kuigi WordNet ja FrameNet on mõlemad leksikaalsed ressursid on nad siiski erinevad, omades mõlemad haid ja halbu külgi. Samas saab neid kasutada koos, kui üks täiustab teist. Hüperonüümia suhted on keeltes enimlevinud, mis võimaldavad kergemalt leida sünonüüme. Seetõttu on WordNetti hea kasutada. Teiselt poolt FrameNetis lisatakse sõnale näitelauseid iga tähenduse ning süntaktilise lausekonstruktsiooni kohta. Näited võetakse naturaalsest korpusest, mitte ei ole leksikograafi poolt välja mõeldud. Tänu sellele on semantilise analüüsi tulemused osutunud paremaks kui WordNetil. Kahe ressursi ühisteks omadusteks on hierarhilised suhted freimide/sünohulkade vahel ning semantiliselt sarnaste sõnade grupeerimine.[4, 24]

Eesti Wordnet põhineb üldisel wordneti teorial. Selles on 100 000 sõna ja 200 000 suhet. Tegemist on Eesti kõige laialdasema leksikaalse andmebaasiga, mis tähendab, et seda on sõnatähenduse määramisel ja muude keeletehnoloogiliste probleemide lahendamisel palju parem kasutada, kui Framenette, milledest üks on valdkonnaspetsiifiline ning teised kaks poolikud.[23]

## 2. Praktilise töö ülevaade

Esimese praktilise ülesandena tuli edasi arendada olemas olnud eestikeelset jalgpalli-alast Framenetti, kuna antud töö hüpoteesi kontrollimiseks on vaja, et igal freimi LÜ-l ning elemendil oleks "morf" atribuut, kus oleks kirjas vastava elemendi kõik võimalikud morfoloogilise ning süntaktilise esinemise võimalused. Teise ülesandena tuli luua programm, mis suudaks morfoloogiliselt ning süntaktiliselt analüüsitud ja ühestatud korpuse tekste automaatselt märgendada freimide ja elementidega, kasutades võrdluseks esimeses ülesandes Framenetti lisatud infot. Viimase osana tuli täiendada Eesti Wordnetti. Loodud ressursid asuvad veebilehel [www.github.com/Tamme/jalgpalliFreimNet](http://www.github.com/Tamme/jalgpalliFreimNet) ning nimekiri failidest on ka lisas 1.

### 2.1. Frameneti täiendamine

Enne automaatse märgendamise juurde asumist tuleb freime täiendada MSI ning lisaatribuutidega, mis hüpoteesi järgi võiks anda piisavalt infot, et täpselt määrata tekstis freime.

Uus Frameneti struktuur on nähtav joonisel 4. Lisandusid atribuut "morf" LÜ-dele ning elementidele, mis hoiab endas MSI-d, "selection", mis on loend sõnadest, mis samuti vastava LÜ või elemendi välja kutsuksid ning "resource", mis täidab samasugust rolli nagu "selection", kuid loend on nii pikk, et seda hoitakse elemendi või LÜ nimelises tekstifailis, kus igal real on üks sõna.

```
<framelexicon>
  <frame name="freimi nimi">
    <LexicalUnits>
      <lu lemma="lemma nimi" selection="seotud mõistete loend" resource="true" morf="morf ja süntaks"/>
      ...
    </LexicalUnits>
    <Elements>
      <element name="nimi" optional="true" selection="loend" resource="true" morf="morf ja süntaks"/>
      ...
    </Elements>
  </frame>
</framelexicon>
```

Joonis 4. Frameneti uus struktuur.

Atribuudid "selection" või "resource" lisati LÜ-dele ja elementidele, mille nimi on hüperonüüm ehk ülemmõiste mingite teiste sõnade suhtes või millel leidub sünonüüme.

Näiteks elemendile nimega "kehaosa" lisati juurde tekstifail, kus on kirjas kõik kehaosad Wordneti järgi, ning elemendile nimega "meeskond" lisati loend, milles sisalduvad sõnad "tiim", "võistkond" ning "klubi".

Põhjused selliseks freimi täiendamiseks oli see, et tekstides kasutatakse samuti sünonüüme ning sõnade täpsemaid alamõisteid, kuid freimide märgendamisel peaks tulemus olema siiski see algsõna.

Ainult MSI abil ei ole varem üritatud eesti keeles teksti freimidega märgendada ning seetõttu tuli reeglistik selle lisamiseks välja mõelda. Reeglistiku loomiseks analüüsiti freime ning mõeldi, millistes lausetes freim võib esineda, millised sõnad võivad selle välja kutsuda ning kuidas elemendid lauses paikneda saavad.

Joonisel 5 on üks valmisolev freim ning sellel on näha ka mõned väljakutsed, mis MSI lisamisel tekkisid. Näiteks on olemas elemendi nimesid, mis on nii abstraktsed, et need võivad esineda ühe sõnana, aga samas olla ka terved fraasid. Sellised on joonisel oleva "finish" freimi puhul "allikas", "käik", "sihtmärk" ning "teekond". Lisaks võib elementide "lööja" ning "sihtmärk" puhul olla tegemist ka pärisnimega. Neid hõlmavaid loendeid kahjuks ei leitud.

Loodud reeglistik on järgnev:

- Kasutada \* esimese sümbolina, kui on fraasina esinemise võimalus, võib olla pärisnimi või on abstraktne ja üldistav sõna, mis võib kontekstis esineda teiste sõnadena.
- Kui ühel elemendil või LÜ-l võib olla mitu morfoloogilist kuju, siis kasutada nende vahel | märki (tähistab loogilist VÕI-d).
- Kui elemendi või LÜ nimi koosneb mitmest sõnast, siis kasutada || sümbolit, mis eraldab sõnade morfoloogilist infot.
- Kui elemendi või LÜ nimi esineb kindlas süntaktilises rollis, siis kasutada morfoloogilise ning süntaktilise info eraldamiseks ||| märki, kusjuures süntaktiline osa peab asuma viimasel kohal.
- Morfoloogilist infot märkida üles nii täpselt kui võimalik, samas kui sõna võib esineda kõikides käänetes, siis võib selle märgendi lihtsalt ära jätta.



- MSI tähistusteks võtta samad, mida kasutati korpuse analüüsil [5].
- Kui lemma või elemendi nimi on mitmesõnaline, siis viia sõnad algkujule.

```
<frame name="finish">
  <LexicalUnits>
    <lu lemma="lõpetama" morf="_V_main" />
    <lu lemma="saatma" morf="_V_main" />
  </LexicalUnits>
  <Elements>
    <element name="lööja" selection="mängija, ründaja, poolkaitsja" optional="true" morf="*_S_||@SUBJ"/>
    <element name="allikas" optional="true" morf="*_S_abl|_S_nom"/>
    <element name="käik" optional="true" morf="*" />
    <element name="kehaosa" resource="true" optional="true" morf="*_S_com" />
    <element name="sihtmärk" optional="true" morf="*_S_all|_S_ill|_S_abl" />
    <element name="teekond" optional="true" morf="*" />
  </Elements>
</frame>
```

Joonis 5. Näidisfreim.

Keeltes esineb sama asja ütlemiseks mitmeid konstruktsioone. Näiteks joonisel 5 oleva freimi puhul on tavaliselt lööja subjekti ehk aluse rollis, kuid võib olla ka käandsõna, "poolt", ees.

Kõikide selliste erijuhtude käsitlemine on sarnane peatükis 1.5.2 kirjeldatud käsitsi loodud grammatikale. Kuna on teada, et selline viis ei toonud edu, siis väga keerulisi lausekonstruktsioone eraldi ei arvestatud ning pigem vaadati, kuidas selline viis töötab lihtsamate lausetega.

MSI-ga freimid asuvad GitHubis failis nimega frames\_with\_morf.xml.

## 2.2. Automaatne märgendamine

Enne automaatse märgendamise juurde asumist tehti seda ühe korpuse tekstiosa puhul käsitsi, kasutades tervikteksti märgendamise meetodit.

Berkeley FrameNeti projekti puhul loodi spetsiaalne kasutajaliides, mille abil märgendamist käsitsi tehakse [25]. Ilma seesuguse programmita on korpuste ja freimide omavaheline võrdlemine üsna aeganõudev ja keeruline, kuna iga sõna puhul tuleb uurida, kas see on mõne freimi jaoks LÜ. Märgendatud tekst asub lisa 2.

Käsitsi märgendamise abil sai selgeks, kui raske on arvutile mingit reeglistikku luua, mille järgi õigesti automaatselt märgendada. Lisada tuli ka uusi LÜ-sid freimidele.

Automaatse märgendamise programm loodi, kasutades PHP (rekursiivne akronüüm nimest PHP: Hypertext Preprocessor) programmeerimiskeelt. Valik langes selle kasuks, kuna

tegemist on kergelt õpitava ning kasutatava skriptikeelega, mis on samas piisavalt võimekas, et selline programm luua [26].

Programmi kasutamiseks peab arvutis leiduma kompileeritud kujul PHP programm (vähemalt versioon 5.3). Seejärel tuleb avada käsuriid, ning liikuda sellega PHP kausta, kus peaks asuma php.exe rakendus. Samal ajal peaks mõne tekstiredaktori abil muutma add-frames.php lähtekoodi algust ning määrama ära konstandid `FRAMES_PATH`, `TEXT_PATH`, `OVERALL_PATH` ning `RETURN_PATH`. Esimene neist peab näitama täisteed freimide failini, teine märgendatava tekstifailini, kolmas üldteed, kus asuvad freimide lisaressursid (freimide sünonüümide ning hüponüümide loendite failid), ning viimane teed failini, kuhu soovitakse kirjutada analüüsitud teksti tulemus. Meeles tuleb pidada, et tee ei tohi sisaldada täpitähti ning sisendina antud failid peavad olema UTF-8 kodeeringus. Seejärel tuleb käsureal sisestada käsk `"php.exe -f täistee add-frames.php failini"` (Windowsi operatsioonisüsteemi puhul näiteks `php.exe -f C:\Kasutajad\Kasutaja\Töölaud\add-frames.php`).

Programmi tööpõhimõtte on võrrelda freimi LÜ-de ning elementide nimesid ning MSI-d korpuse analüüsitud tekstidega.

Selleks, et rakendus suudaks seda efektiivselt teha, tuli freimide XML fail ning korpuse tekstifailid esmalt massiividenä esitada. XML on kindla formaadiga standard, kuid korpuse tekstifailid ei olnud piisavalt ühtse struktuuriga, et neid oleks hea lugeda, seega seati nendele järgmised nõuded:

- sõna algvormi ees peab olema täpselt 4 tühikut
- MSI märgendite ümber peavad olema tühikud
- sõna, kõik tema algvormid ja morfoloogiline info peab olema ühel real

Nende nõuete järel on võimalik ka tekstifailid massiivide kujule viia. Loodud massiivide üldkujud on näha joonistel 6 ja 7, kus ülakomade vahel asetsevad sõnad on fikseeritud ning ülejäänud muutuvad või on lihtsalt loendurid.

```

laused = (
  [lauseLoendur] => (
    [sõnaLoendur] => (
      ['alg'] (
        [algvormiLoendur] => sõna algvorm
        ...
      ),
      ['morf'] (
        [algvormiLoendur] => algvormi MSI
        ...
      ),
      'nr' => mitmes on see sõna tekstis
    ),
    ...
  ),
  ...
)

```

Joonis 6. Lausete massiiv.

```

frames = (
  [freimiNimi] => (
    ['lexicalUnits'] => (
      [leksikaalseÜksuseNimi] => (
        ['morf'] => leksikaalse üksuse morfoloogiline info
        [] => leksikaalse üksuse nimi
        [] => leksikaalsele üksusele vastavate ressursside nimed, igaüks eraldi liikmena
        ...
      ),
      ...
    ),
    ['elements'] => (
      [elemendiNimi] => (
        ['morf'] => elemendi morfoloogiline info
        [] => elemendi nimi
        [] => elemendile vastavate ressursside nimed, igaüks eraldi liikmena
        ...
      ),
      ...
    )
  ),
  ...
)

```

Joonis 7. Freimide massiiv.

Programmi algoritm on üsna lihtne, kuna freimide MSI ning ressursside määramisel peaks olema kõikide võimaluste otsustamine tehtud ning jääb ainult freimi- ning tekstimassiivi omavaheline võrdlemine lähtudes varem seatud reeglitest. Täpsem eeskiri on nähtav joonisel 8, mis üldistatult analüüsib tekstimassiivi igat lauset iga freimiga kõikide LÜ-dega ning sobivuse korral käsitleb lauset uuesti sama freimi kõikide elementidega. Analüüsi osa toimub vastavalt algoritmile, mis järgib pseudokoodi joonisel 9.

```

for iga lause L tekstimassiivis do
  for iga freim F freimimassiivis do
    for iga leksikaalne üksus LÜ freimis F do
      analüüsi lauset L freimi F leksikaalse üksusega LÜ
      if analüüsi käigus lisati lausele freim F then
        for iga element E freimis F do
          analüüsi lauset L freimi F elemendiga E
        end
      end
    end
  end
end

```

Joonis 8. Freimi- ning tekstimassiivi võrdlemine.

```

for LÜ/E iga nimi või ressurss do
  if E/LÜ morf algab tärniga then
    arvesta seda ning eemalda morfish
  end
  if E/LÜ morf sisaldab || then
    poolita selle juurest, määra esimene osa morfish ning teine osa süntaksiks
  end
  if E/LÜ morf sisaldab || then
    tükelda nende kohtade juurest
    tükelda nimi või ressurss
  end
  for iga morf tükk do
    if tükk sisaldab | then
      võimalus = tükelda tükk |-de juurest
    end
  end
  for iga tükk do
    for sõna S lauses L do
      if sõnal peaks olema süntaks AND süntaks ei vasta sõna morfile then
        continue
      end
      for sõna S morfi tüki võimalus do
        if morf algas tärniga then
          if kõik E/LÜ morfi tüki võimaluse osad asuvad S morfi sees AND sõna = E/LÜ nimega then
            lisa tagastatava teksti rea lõppu '// Freim/Element E/freimi nimi'
          else if sõna = E/LÜ then
            lisa tagastatava teksti rea lõppu '// Freim/Element õnnega2 E/freimi nimi'
          else if kõik E/LÜ morfi tüki võimaluse osad asuvad S morfi sees then
            lisa tagastatava teksti rea lõppu '// Freim/Element õnnega1 E/freimi nimi'
          end
        else
          if kõik E/LÜ morfi tüki võimaluse osad asuvad S morfi sees AND sõna = E/LÜ nimega then
            lisa tagastatava teksti rea lõppu '// Freim/Element E/freimi nimi'
          end
        end
      end
    end
  end
end

```

Joonis 9. Lause ning LÜ või elemendi analüüsi pseudokood.

Nagu jooniselt 9 on näha, kirjutatakse tagastatavasse faili freime ning elemente kolmel viisil:

- // Freim/Element *elemendi/LÜ nimi*, kui freimi MSI sisaldub sõna MSI-s ja freimi elemendi/LÜ nimi ning sõna langevad kokku
- // Freim/Element õnnega1 *elemendi/LÜ nimi*, kui freimi morf algab tärniga ning freimi MSI sisaldub sõna MSI-s
- // Freim/Element õnnega2 *elemendi/LÜ nimi*, kui freimi morf algab tärniga ning freimi elemendi/LÜ nimi ning sõna langevad kokku

Kuju ”// Freim/Element (õnnega2) *elemendi/LÜ nimi*” tähendab seda, et kui esimene kord analüüsitakse freimi LÜ-sid, siis kirjutatakse ”// Freim (õnnega1/2) *freimi nimi*” ning kui sobiva freimi korral hakatakse elemente vaatama, siis kirjutatakse ”// Element (õnnega1/2) *elemendi nimi*”. Ühes reas võib olla ka mitu elemendi või freimi märgendit, kui MSI ning LÜ või elemendi nime põhjal on märgendamiseks mitu võimalust.

Esimese variandi puhul võib tulemuse rikkuda vaid sõna mitmetähenduslikkus, mida selles valdkonnaspetsiifilises Framenetis esines üsna vähe. Teised kaks on heuristilised ning nende täpsus sõltub korpusest ning freimidele lisatud MSI spetsiifilisusest.

Joonisel 10 on üks automaatselt märgendatud lause. Sellelt on näha, et väga palju on heuristiliselt valitud freime ning elemente, mis ei ole täppi läinud. Lisaks on ühe sõna puhul määratud mitu freimi, mis tähendab, et see sõna esineb mitme freimi LÜ-na. See ei ole ainult ühe lause tulemus, vaid kogu tekst märgendamise tulemus on sarnane. Seetõttu ei saa hüpoteesi õigeks pidada ning MSI-st ei piisa automaatseks märgendamiseks. Tavaliselt määrati õigesti vaid need sõnad, mis esinesid Framenetis ning ei kujutanud midagi üldist või abstraktset.

```
$LAS ##### @???  
Kaks kaks+0 // _N_ card sg nom #cap 1 // **CLB @SUBJ @ADVL  
minutit minut+0 // _S_ com sg part // @<Q // Freim õnnega1 goal // Element õnnega1 kehaosa // Element õnnega1 sihtmärk...  
hiljem hiljem+0 // _D_ // @ADVL  
teenis teenis+0 // _V_ main indic impf ps3 sg ps af // Freim concede compensation // Freim receive card  
Voskoboinikov Voskoboinikov+0 // _S_ prop sg nom #cap #? // @SUBJ @NN> // Freim õnnega1 goal // Element õnnega1 lööja...  
kollase kollane+0 // _A_ pos sg gen // @AN> // Freim sanction  
kaardi kaart+0 // _S_ com sg gen // @ADVL // Freim õnnega1 goal // Element õnnega1 kehaosa // Element õnnega1 sihtmärk...  
$. // _Z_ Fst //  
$LLS ##### @???
```

Joonis 10. Märgendamise tulemuse näidis.

Antud metoodikat ei ole mõtet edasi arendada, kuna masinõppele sarnast täpsust (umbes 65%) on sellisel viisil väga raske saavutada. Võimalus on tööd jätkata eestikeelse üldise

Framenetiga, kuid seal oleks vaja teha väga palju käsitsi märgendamist. Uuritud on ka Wordneti ning Frameneti integreerimise võimalusi [27]. Sellisel juhul Framenet rikastaks Wordnetti freimide semantilise informatsiooniga, mis võiks parandada tekstide tähenduse määramist Wordneti abil. Kuna eesti keele Wordnet on piisavalt mahukas, siis võiks järgmisena uurida võimalusi Wordneti ning Frameneti sidumiseks.

## **2.3. Eesti Wordneti täiendamine**

Maailmas alustati Wordnettide loomist varem kui freimidel põhinevaid ressursse ning seetõttu on ka eesti keeles kõige mahukama leksikoni puhul tegemist Wordnetiga [23]. Sellest tulenevalt tasub seda alati täiendada, kui leidub sõnu, mis seal veel ei ole.

Viimase praktilise ülesandena analüüsiti läbi kõik kasutatud korpused ja freimid ning vaadati, kas leidub sõnu, mida Eesti Wordnetis ei ole või mis omab mõnda teistsugust tähendust. Selle tarbeks kasutati tervet Wordnetti hõlmavat tekstifaili ning iga jalgpalli-alase mõiste puhul otsiti, kas sõna asub seal ning kas tal on vastav tähendus.

Kokku õnnestus leida 183 sõna või väljendit, mida võib kindlalt sinna lisada ning 56 väljendit, mille puhul tuleb otsustada, kas need sobivad sinna või mitte. Viimaste hulgas on palju ühend- ning väljendverbe, näiteks "skoori avama", millede puhul on keeruline öelda, kas need moodustavad koos väljendi või on tegemist lihtsalt kahe omaette sõnaga.

Selles töö osas valmis vaid nimekiri leitud sõnadest, mis antakse edasi igapäevaselt Wordnetiga tegelevatele inimestele, kes ise otsustavad, kui suur osa sõnadest tuleks lisada.

Nimekiri uutest sõnadest koos tähendusega asub lisas 3 ning GitHubis, failis toWordnet.txt

## Kokkuvõte

Juba üle poole sajandi on loomuliku keele analüüs arvutite abil olnud ülesandeks, mida üritatakse lahendada. Raskeim osa sellest on tekstide semantika mõistmine.

Freimiteooria on üks viis, kuidas seda on üritatud lahendada. Sellega seonduvat tekstide automaatset märgendamist freimidega selles töös uuritigi.

Suurim freimide kogum on ingliskeelne FrameNet, kus on üle 10 000 LÜ ning 150 000 näidislause, mis annab võimaluse kasutada masinõpet nende lausete märgendamiseks, mis ei esine märgendatud näidislausete hulgas.

Eesti keele üldine Framenet on alles algusjärgus, kuid on ka jalgpalli-alane freimide ressurss, mis on terviklikum. Kasutades teisena nimetatud resurssi, üritati selles töös uurida, kas morfoloogiliselt ning süntaktiliselt analüüsitud tekstist ja MSI-ga freimidest piisab, et vastava valdkonna tekste saaks edukalt märgendada.

Loodud programmi tulemus näitab kahjuks, et selline info ei ole piisav tekstide märgendamiseks ning õiged tulemused saadi peamiselt vaid juhul, kui LÜ või elemendi nimi asus tekstis. Kui tegemist oli üldise sõnaga nagu näiteks "lööja" või "allikas", siis MSI ei piiranud valikut piisavalt, et märgendamine oleks edukas.

Metoodika võib natukene paraneda täpsema MSI kirjeldamisel ning Frameneti suurendamisel, kuid otstarbekam oleks uurimist jätkata Frameneti ning Wordneti kombineerimise teel, kuna eestikeelne Wordnet on juba nii mahukas, et võimaldab teha üldtekstilisi analüüse.

Uurimistööd võib siiski pidada edukaks, kuna ka teadmine, et sellisel viisil märgendamine ei õnnestu, on kasulik info ning sai loodud fail, milles on kirjas jalgpalli-alased mõisted, mis Eesti Wordnetis ei esine, kuid mis loodetavasti lisatakse sinna peagi.

## Kasutatud kirjandus

- [1] M. Scaiano, D. Inkpen (2011), "Machine Learning for Automatic Labeling of Frames and Frame Elements in Text", *International Journal of Linguistics*, Vol. 3(1), [http://www.macrothink.org/journal/index.php/ijl/article/download/648/pdf\\_12](http://www.macrothink.org/journal/index.php/ijl/article/download/648/pdf_12)
- [2] C. J. Fillmore (1976), "Frame semantics and the nature of language", *Annals of the New York Academy of Sciences*, Vol. 280, <http://www.icsi.berkeley.edu/pubs/ai/frame semantics76.pdf>
- [3] D. Gildea, D. Jurafsky (2002), "Automatic Labeling of Semantic Roles", *Computational Linguistics*, Vol. 28(3), <http://www.cs.rochester.edu/~gildea/gildea-cl02.pdf>
- [4] J. Ruppenhofer, M. Ellsworth, M. R. L. Petruck et al. (2010), "FrameNet II: Extended Theory and Practice", <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>
- [5] A. Orlo (2014), "Jalgpalli Frameneti tõlkimine eesti keelde ja jalgpalli tekstikorpuse koostamine", *Tartu Ülikooli bakalaureusetöö*
- [6] H. Õim, Freimileksikon, <http://www.keeletehnoloogia.ee/ekt-projektid/semantika-vahendid-eesti-keelele/freimileksikon>, vaadatud 12.05.14
- [7] Rätsepa korpus, <http://www.keeletehnoloogia.ee/projektid/semantika/ressursid/ratsepakorpus-tiger.xml>, vaadatud 13.05.14
- [8] Freimileksikon, <http://www.keeletehnoloogia.ee/projektid/semantika/ressursid/freimileksikon>, vaadatud 13.05.14
- [9] About FrameNet, <https://framenet.icsi.berkeley.edu/fndrupal/about>, vaadatud 12.05.14
- [10] D. Das, D. Chen, A. F. T. Martinsy, et al. (2014), "Frame-semantic Parsing", *Computational Linguistics*, Vol. 40(1) <http://static.googleusercontent.com/media/research.google.com/et/pubs/archive/41227.pdf>
- [11] XML Essentials, <http://www.w3.org/standards/xml/core>, vaadatud 08.05.14
- [12] Analüüsi etapid, <http://math.ut.ee/~kaili/parser/demo/>, vaadatud 08.05.14
- [13] ESTKG süntaktilised märgendid, [http://math.ut.ee/~kaili/thesis/pt3\\_4.html](http://math.ut.ee/~kaili/thesis/pt3_4.html), vaadatud 08.05.14
- [14] Ülevaade morfoloogilistest märgenditest, [http://math.ut.ee/~kaili/thesis/pt3\\_2.html](http://math.ut.ee/~kaili/thesis/pt3_2.html), vaadatud 08.05.14



- [15] Morfosüntaktilised kategooriad, <http://www.cl.ut.ee/korpused/morfliides/seletus>, vaadatud 12.05.14
- [16] C. A. Thompson, R. Levy, C. D. Manning (2003), "A Generative Model for Semantic Role Labeling", *Machine Learning: ECML 2003*, [http://link.springer.com/chapter/10.1007%2F978-3-540-39857-8\\_36](http://link.springer.com/chapter/10.1007%2F978-3-540-39857-8_36)
- [17] M. Ereht, T. Ereht, E. Veldi (2012), "Eesti-inglise keeleteaduse sõnastik"
- [18] D. Das, N. Schneider, D. Chen, et al. (2010), "SEMAFOR 1.0: A Probabilistic Frame-Semantic Parser", *HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, <http://www.ark.cs.cmu.edu/SEMAFOR/das+schneider+chen+smith.tr10.pdf>
- [19] D. Das, N. Schneider, D. Chen, et al. (2010), "Probabilistic Frame-Semantic Parsing", *HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, <http://www.cs.cmu.edu/~nasmith/papers/das+schneider+chen+smith.naacl10.pdf>
- [20] M. Fleischman, E. Hovy (2003), "A Maximum Entropy Approach to FrameNet Tagging", *NAACL-Short '03 Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers - Volume 2*, <http://acl.ldc.upenn.edu/N/N03/N03-2008.pdf>,
- [21] D. Das, A. F. T. Martins, N. A. Smith (2012), "An Exact Dual Decomposition Algorithm for Shallow Semantic Parsing with Constraints", *SemEval '12 Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, [http://www.cs.cmu.edu/~afm/Home\\_files/starsem2012.pdf](http://www.cs.cmu.edu/~afm/Home_files/starsem2012.pdf)
- [22] What is WordNet?, <http://wordnet.princeton.edu/>, vaadatud 12.05.14
- [23] TEKSaurus, <http://www.cl.ut.ee/ressursid/teksaurus/index.php?lang=et>, vaadatud 12.05.14
- [24] J. N. Zheng, L. Melchor, F. Omur (2008), WordNet & FrameNet, <http://www.umiacs.umd.edu/~jimmylin/INFM700-2008-Spring/presentations/WordNet.ppt>
- [25] C. J. Fillmore, C. F. Baker, H. Sato (2002), "The FrameNet Database and Software Tools", *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, <http://www.icsi.berkeley.edu/pubs/ai/framenetdatabase02.pdf>
- [26] What is PHP?, <http://www.php.net/manual/en/intro-what-is.php>, vaadatud 08.05.14

[27] E. Laparra, G. Rigau (2009), "Integrating WordNet and FrameNet using a knowledge-based Word Sense Disambiguation algorithm", *Recent Advances in Natural Language Processing*, <http://adimen.si.ehu.es/~rigau/publications/ranlp09-lr.pdf>

# Lisad

## Lisa 1

Antud töös loodud failid on järgmised:

- add-frames.php – programm, mis märgendab automaatselt
- frames\_with\_morf.xml – MSI-ga jalgpalli-alane Framenet
- toWordnet.txt – loend sõnadest ning mõistetest, mida võiks lisada Eesti Wordnetti
- addFramesREADME.txt – juhend, kuidas programmi kasutada
- soccernet\_manual\_tag.snx – üks korpuse tekst, mis on osaliselt käsitsi märgendatud

## Lisa 2

### Käsitsi märgendatud tekst

\$LAS\$ ##### \*\*CLB @???

Juventus Juventus+0 // \_S\_ prop sg nom #cap #? // \*\*CLB @SUBJ //Element võitja

tegi tege+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV // Freim victory

Udinesele Udinese+le // \_S\_ prop sg all #cap #? // @ADVL Udines+le // \_S\_ prop sg all #cap #? // @ADVL //Element kaotaja

kiirelt uuesti+0 // \_D\_ // @ADVL

otsad võistlus+d // \_S\_ com pl nom // @OBJ @NN> // Freim victory

\$LL\$ ##### @???

\$LAS\$ ##### @???

Itaalia Itaalia+0 // \_S\_ prop sg nom #cap // @NN>

Serie Serie+0 // \_S\_ prop sg nom #cap #? // @<NN

A-s A+s // \_Y\_ sg in #cap #? // @ADVL @<NN

kinnitas alusta+s // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV //Freim lead

oma oma+0 // \_P\_ pos det refl sg gen // @NN>

eduseisu edu\_seis+0 // \_S\_ com sg gen // @OBJ @NN>

Juventus 5:0+0 // \_N\_ prop sg nom #cap #? // @<NN //Element liider

\$, , // \_Z\_ Com //

kes kes+0 // \_P\_ inter rel sg nom // \*\*CLB @SUBJ //Element võitja

oli ole+i // \_V\_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV //Freim victory

võõral võõras+l // \_A\_ pos sg ad // @AN> //Element mängu asukoht

väljakul väljak+l // \_S\_ com sg ad // @ADVL //Element mängu asukoht

üle üle+0 // \_D\_ // @ADVL //Freim victory

Udinesele Udinese+st // \_S\_ prop sg el #cap #? // @ADVL @<NN Udines+st // \_S\_ prop sg el #cap #? // @ADVL @<NN //Element kaotaja

\$. . // \_Z\_ Fst //

\$LL\$ ##### @???

\$LAS\$ ##### @???

Juventus Juventus+0 // \_S\_ prop sg gen #cap #? // \*\*CLB @NN> //Element meeskond // Element võitja

treener treener+0 // \_S\_ com sg nom // @NN> //Element treener

Antonio Antonio+0 // \_S\_ prop sg nom #cap // @NN> //Element treener

Conte Conte+0 // \_S\_ prop sg nom #cap #? // @SUBJ //Element treener

otsustas otsusta+s // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P #InfP // @+FMV

esmaspäeva esmas\_päev+0 // \_S\_ com sg gen // @OBJ @NN>

õhtul õhtu+l // \_S\_ com sg ad // @ADVL  
 toimunud toimunud+0 // \_A\_ pos #nud // @VN>  
 mängus mäng+s // \_S\_ com sg in // @ADVL @NN> //Element mäng  
 puhkust puhkus+t // \_S\_ com sg part // @OBJ //Freim deploy  
 anda anda+a // \_V\_ main inf #NGP-P #InfP #all #tr // @OBJ  
 mitmele mitu+le // \_P\_ inter rel indef sg all // @ADVL //Element mitte mänigja  
 põhitegijale põhi tegija+le // \_S\_ com sg all // @<Q //Element mitte mänigja  
 \$, , // \_Z\_ Com //  
 ent ent+0 // \_J\_ crd // \*\*CLB @J  
 sellele see+le // \_P\_ dem sg all // @P>  
 vaatamata vaatamata+0 // \_D\_ // @ADVL vaatamata+0 // \_K\_ post #all // @ADVL vaata+mata // \_V\_ main sup ps abes #NGP-P // @ADVL  
 tuldi tule+di // \_V\_ main indic impfimps af #FinV #Intr // @+FMV //Freim victory  
 Udinesega Udinese+ga // \_S\_ prop sg kom #cap #? // @ADVL @NN> Udines+ga // \_S\_ prop sg kom #cap #? // @ADVL @NN> //Element kaotaja  
 toime toim+0 // \_S\_ com sg adit // @ADVL @<NN toime+0 // \_S\_ com sg adit // @ADVL @<NN //Freim victory  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 Otsustavaks otsustav+ks // \_A\_ pos sg tr #cap // \*\*CLB @ADVL // Element otsustav sündmus  
 sai saa+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 Juventuse Juventus+0 // \_S\_ prop sg gen #cap #? // @NN> // Element võitja  
 kvaliteet kvaliteet+0 // \_S\_ com sg nom // @SUBJ // Element otsustav sündmus  
 rünnakul rünnak+l // \_S\_ com sg ad // @ADVL @<NN // Element otsustav sündmus  
 ning ning+0 // \_J\_ crd // \*\*CLB @J  
 mäng mäng+0 // \_S\_ com sg nom // @OBJ // Element mäng  
 kindlustati kindlusta+ti // \_V\_ main indic impfimps af #FinV #NGP-P // @+FMV // Freim victory  
 ära ära+0 // \_D\_ // @ADVL // Freim victory  
 vähem vähem+0 // \_A\_ comp sg nom // @AN> vähem+0 // \_D\_ // @ADVL  
 kui kui+0 // \_J\_ crd sub // @J  
 poole pool+0 // \_N\_ card sg gen l // @ADVL  
 tunniga tund+ga // \_S\_ com sg kom // @<Q  
 tänu tänu+0 // \_K\_ pre #all // @ADVL tänu+0 // \_S\_ com sg gen // @ADVL @NN> tänu+0 // \_S\_ com sg nom // @NN> @ADVL tänu+0 // \_S\_ com sg part // @SUBJ @ADVL  
 Sebastian Sebastian+0 // \_S\_ prop sg nom #cap #? // @NN>  
 Giovinco Giovinco+0 // \_S\_ prop sg nom #cap #? // @SUBJ @NN> @ADVL  
 ja ja+0 // \_J\_ crd // \*\*CLB-C @J

Fernando Fernando+0 // \_S\_ prop sg nom #cap #? // @NN>  
 Llorente Llorente+0 // \_S\_ prop sg gen #cap #? // @ADVL @NN>  
 väravatele värav+tele // \_S\_ com pl all // @<P @ADVL @<NN  
 \$. // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 PSG PSG+0 // \_Y\_ ? #cap #? // \*\*CLB @SUBJ @ADVL // Element kaotaja  
 teenis teeni+s // \_V\_ main indic impf ps3 sg ps af #FinV #Part-P // @+FMV  
 hooaja hoo\_aeg+0 // \_S\_ com sg gen // @NN>  
 teise teine+0 // \_P\_ indef dem sg gen // @NN>  
 kaotuse kaotus+0 // \_S\_ com sg gen // @NN> // Freim defeat  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Pariisi Pariis+0 // \_S\_ prop sg gen #cap // @NN>  
 suurlubi suur\_klubi+0 // \_S\_ com sg gen // @NN>  
 Saint-Germain Saint-Germain+0 // \_S\_ prop sg nom #cap #? // @SUBJ @NN> // Element kaotaja  
 sai sai+0 // \_S\_ com sg nom // @SUBJ @NN>  
 Ligue Ligue+0 // \_S\_ prop sg gen #cap #? // @NN> // Element võistlus  
 1 1+0 // \_N\_ card ? digit // @SUBJ @ADVL @NN> @<NN // Element võistlus  
 tabelisse tabel+sse // \_S\_ com sg ill // @ADVL @<Q @NN> @<NN  
 kirja kiri+0 // \_S\_ com sg gen // @NN> @<Q  
 hooaja hoo\_aeg+0 // \_S\_ com sg gen // @NN> @<Q  
 teise teine+0 // \_N\_ ord sg gen 1 // @AN>  
 kaotuse kaotus+0 // \_S\_ com sg gen // @ADVL @<Q // Freim defeat  
 \$, , // \_Z\_ Com //  
 kui kui+0 // \_J\_ crd sub // \*\*CLB @J  
 võõral võõras+1 // \_A\_ pos sg ad // @AN> // Element väljak  
 väljakul väljak+1 // \_S\_ com sg ad // @ADVL // Element väljak  
 jäädi jää+di // \_V\_ main indic impfimps af #FinV #Intr // @+FMV  
 0:1 0:1+0 // \_N\_ card ? digit // @ADVL @NN> // Element lõppskoor  
 alla alla+0 // \_K\_ pre #gen // @ADVL  
 Lyoni Lyon+0 // \_S\_ prop sg gen #cap // @<P // Element võitja  
 meeskonnale meeskond+le // \_S\_ com sg all // @ADVL @<Q // Element võitja  
 \$. // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???

ilma ilma+0 // \_D\_ #cap // \*\*CLB @ADVL ilm+0 // \_S\_ com sg adit #cap // \*\*CLB @ADVL @NN>  
 ilm+0 // \_S\_ com sg gen #cap // \*\*CLB @NN>  
 Zlatan Zlatan+0 // \_S\_ prop sg nom #cap #? // @NN>  
 Ibrahimovicit Ibrahimovicit+0 // \_S\_ prop sg nom #cap #? // @NN>  
 mängival mängi=v+1 // \_A\_ pos sg ad #v partic // @VN>  
 PSG1 PSG+1 // \_S\_ prop sg ad #cap #? // @ADVL @<NN // Element kaotaja // Element meeskond  
 ei ei+0 // \_V\_ aux neg // @NEG // Freim miss chance  
 õnnestunud õnnestu=nud // \_V\_ main indic impf ps neg #FinV // @+FMV // Freim miss chance  
 esimest esimene+t // \_N\_ ord sg part 1 // @AN> esimene+t // \_P\_ dem sg part // @NN>  
 korda kord+0 // \_S\_ com sg part // @OBJ @ADVL  
 tänavause tänavaus+0 // \_S\_ com sg gen #? // @NN>  
 hooaja hooaeg+0 // \_S\_ com sg gen // @P>  
 jooksul jooksul+0 // \_K\_ post #gen // @ADVL  
 väravani värav+ni // \_S\_ com sg term // @ADVL  
 jõuda jõud+a // \_V\_ main inf #InfP #ill #all #ter // @SUBJ  
 ning ning+0 // \_J\_ crd // \*\*CLB @J  
 võeti võt+ti // \_V\_ main indic impfimps af #FinV #NGP-P // @+FMV  
 vastu vastu+0 // \_D\_ // @ADVL  
 teine teine+0 // \_N\_ ord sg nom 1 // @AN>  
 kaotus kaotus+0 // \_S\_ com sg nom // @OBJ @NN> @ADVL // Freim defeat  
 liigas liiga+s // \_S\_ com sg in // @ADVL @<NN // Element võistlus  
 \$. // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Mitmeid mitu+id // \_P\_ inter rel indef pl part #cap // \*\*CLB @NN>  
 häid hea+id // \_A\_ pos pl part // @AN>  
 võimalusi võimalus+i // \_S\_ com pl part // @OBJ // Element võimalus  
 raisanud raisa=nud+0 // \_A\_ pos #nud partic // @VN> // Freim miss chance  
 pariislattel pariislane+tel // \_S\_ com pl ad // @ADVL @<NN  
 jättis jät+is // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P #ill #all #tr // @+FMV  
 soovida soovi+da // \_V\_ main inf #Part #InfP // @-FMV  
 eelkõige eelkõige+0 // \_D\_ // @ADVL  
 otsutavus otsu=tavus+0 // \_S\_ com sg nom #tavus #? // @SUBJ  
 rünnakute rünnak+te // \_S\_ com pl gen // @NN>  
 lõpetamisel lõpetamine+l // \_S\_ com sg ad // @ADVL @<NN  
 \$. // \_Z\_ Fst //

\$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Mängu mäng+0 // \_S\_ com sg gen #cap // \*\*CLB @NN>  
 ainus ainus+0 // \_A\_ pos sg nom // @AN>  
 värav värav+0 // \_S\_ com sg nom // @SUBJ // Freim score goal  
 sündis lõpeta+s // \_V\_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV  
 31. 31.+0 // \_N\_ ord ? digit // @AN>  
 minutil minut+1 // \_S\_ com sg ad // @ADVL  
 Jordan Jordan+0 // \_S\_ prop sg nom #cap #? // @NN> // Element lööja  
 Ferri Ferri+0 // \_S\_ prop sg gen #cap #? // @NN> Ferr+0 // \_S\_ prop sg gen #cap #? // @NN> // Element lööja  
 25'lt 25+lt // \_S\_ com sg abl #? // @ADVL @<NN  
 meetrilt meeter+lt // \_S\_ com sg abl // @ADVL @<NN  
 sooritatud soorita=tud+0 // \_A\_ pos #tud partic // @VN>  
 pealelöögist peale\_löök+st // \_S\_ com sg el // @ADVL // Element löök  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 PSG PSG+0 // \_Y\_ ? #cap #? // \*\*CLB @SUBJ @OBJ @ADVL  
 päästis pääst+is // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 suuremast suure=m+st // \_A\_ comp sg el #m // @AN>  
 kaotusest kaotus+st // \_S\_ com sg el // @ADVL @NN>  
 Salvatore Salvatore+0 // \_S\_ prop sg nom #cap #? // @NN> // Element väravavaht  
 Sirigu Sirigu+0 // \_S\_ prop sg nom #cap #? // @SUBJ @<NN // Element väravavaht  
 \$, , // \_Z\_ Com //  
 kes kes+0 // \_P\_ inter rel sg nom // \*\*CLB @SUBJ  
 suutis suut+is // \_V\_ main indic impf ps3 sg ps af #FinV #InfP // @+FMV  
 mitmel mitu+1 // \_P\_ inter rel indef sg ad // @ADVL @NN>  
 korral kord+1 // \_S\_ com sg ad // @<Q  
 näidata näita+ta // \_V\_ main inf #Part-P // @OBJ  
 suurepärast suure\_pärane+t // \_A\_ pos sg part // @AN>  
 tõrjetööd tõrje\_töö+d // \_S\_ com sg part // @OBJ // Freim save  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Võidu võidu+0 // \_D\_ #cap // \*\*CLB @ADVL võit+0 // \_S\_ com sg gen #cap // \*\*CLB @ADVL // Freim victory



teenis teeni+s // \_V\_ main indic impf ps3 sg ps af #FinV #Part-P // @+FMV  
 liiga liiga+0 // \_D\_ // @ADVL liiga+0 // \_S\_ com sg gen // @NN> liig+0 // \_S\_ com sg adit // @ADVL @NN> liig+0 // \_S\_ com sg part // @OBJ // Element võistlus  
 teine teine+0 // \_N\_ ord sg nom 1 // @AN>  
 meeskond meeskond+0 // \_S\_ com sg nom // @SUBJ @ADVL  
 \$, , // \_Z\_ Com //  
 Monaco Monaco+0 // \_S\_ prop sg nom #cap // @SUBJ @<NN // Element võitja  
 \$, , // \_Z\_ Com //  
 kes kes+0 // \_P\_ inter rel sg nom // \*\*CLB @ADVL  
 Emmanuel Emmanuel+0 // \_S\_ prop sg nom #cap #? // @SUBJ @NN> // Element otsustav sündmus  
 Riviere rivi\_ere+0 // \_A\_ pos sg nom #cap // @ADVL // Element otsustav sündmus  
 väravast värav+st // \_S\_ com sg el // @ADVL @<NN // Element otsustav sündmus  
 alistas alista+s // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 Rennes'i Rennes+i // \_S\_ prop sg part #cap #? // @OBJ /7 Element kaotaja  
 1:0 1:0+0 // \_N\_ card ? digit // @SUBJ @ADVL // Element lõppskoor  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 Monaco Monaco+0 // \_S\_ prop sg nom #cap // \*\*CLB @SUBJ  
 vähendas vähenda+s // \_V\_ main indic impf ps3 sg ps af #FinV #Part // @+FMV  
 vahe vahe+0 // \_S\_ com sg gen // @P>  
 peale peale+0 // \_K\_ post #gen // @ADVL  
 33. 33.+0 // \_N\_ ord ? digit // @AN>  
 vooru voor+0 // \_S\_ com sg gen // @ADVL @NN>  
 PSG'ga PSG+ga // \_S\_ prop sg kom #cap #? // @ADVL @NN> @<NN  
 kümnele kümme+le // \_N\_ card sg all 1 // @ADVL @<NN  
 punktile punkt+le // \_S\_ com sg all // @<Q  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 Lille lill+0 // \_S\_ com sg adit #cap // \*\*CLB @ADVL lill+0 // \_S\_ com sg gen #cap // \*\*CLB @OBJ  
 kindlustas kindlusta+s // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 enda ise+0 // \_P\_ pos det refl sg gen // @NN>  
 kolmandat kolmas+t // \_N\_ ord sg part 1 // @AN>  
 positsiooni positsioon+0 // \_S\_ com sg part // @OBJ @ADVL  
 \$, , // \_Z\_ Com //

edestades edesta+des // \_V\_ main ger af #Part // \*\*CLB-C @ADVL  
 tabeli tabel+0 // \_S\_ com sg gen // @NN>  
 neljandat neljas+t // \_N\_ ord sg part 1 // @AN>  
 Saint-Etienne'i Saint-Etienne+i // \_S\_ prop sg part #cap #? // @OBJ  
 seitse seitse+0 // \_N\_ card sg gen 1 // @ADVL  
 silmaga silm+ga // \_S\_ com sg kom // @<Q  
 \$. // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Vaštšuk Vaštšuk+0 // \_S\_ prop sg nom #cap #? // \*\*CLB @SUBJ  
 lõbutses lõbutse+s // \_V\_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV  
 Blackburni Blackburni+0 // \_S\_ prop sg gen #cap #? // @NN> Blackburn+0 // \_S\_ prop sg adit #cap #? // @ADVL Blackburn+0 // \_S\_ prop sg gen #cap #? // @NN>  
 vastu vastu+0 // \_D\_ // @ADVL  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Readingu Readingu+0 // \_S\_ prop sg gen #cap #? // @NN> Reading+0 // \_S\_ prop sg gen #cap #? // @NN>  
 noorteakadeemiat noor+te\_akadeemia+t // \_S\_ com sg part // @OBJ  
 esindav esinda=v+0 // \_A\_ pos sg nom #v partic // @VN>  
 Bogdan Bogdan+0 // \_S\_ prop sg nom #cap #? // @NN> // Element lõõja  
 Vaštšuk Vaštšuk+0 // \_S\_ prop sg nom #cap #? // @<NN // Element lõõja  
 võttis võt+is // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 ainuisikuliselt ainu\_isiku=liselt+0 // \_D\_ // @ADVL  
 juppideks jupp+deks // \_S\_ com pl tr // @ADVL  
 Blackburni Blackburni+0 // \_S\_ prop sg gen #cap #? // @NN> Blackburn+0 // \_S\_ prop sg gen #cap #? // @NN>  
 eakaaslased ea\_kaaslane+d // \_S\_ com pl nom // @OBJ  
 \$, // \_Z\_ Com //  
 mängides mängi+des // \_V\_ main ger af #Part-P // \*\*CLB @ADVL  
 3:1 3:1+0 // \_N\_ card ? digit // @SUBJ @OBJ @ADVL  
 võidus võidu+s // \_V\_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV  
 rolli roll+0 // \_S\_ com sg adit // @ADVL @NN> roll+0 // \_S\_ com sg gen // @OBJ @NN>  
 kahe kaks+0 // \_N\_ card sg gen 1 // @P>  
 värava värav+0 // \_S\_ com sg gen // @<Q // Freim score goal // Element värav  
 juures juures+0 // \_K\_ post #gen // @ADVL  
 \$. // \_Z\_ Fst //

\$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 \$" " // \_Z\_ Quo //  
 3:1 3:1+0 // \_N\_ card ? digit // \*\*CLB @ADVL // Element lõppskoor  
 võitsime võit+sime // \_V\_ main indic impf ps1 pl ps af #FinV #NGP-P #in // @+FMV // Freim victory  
 Blackburni Blackburn+0 // \_S\_ prop sg part #cap #? // @OBJ // Element kaotaja  
 \$, , // \_Z\_ Com //  
 täna täna+0 // \_D\_ // \*\*CLB @ADVL  
 oli ole+i // \_V\_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV  
 minu mina+0 // \_P\_ pers ps1 sg gen // @NN>  
 päev päev+0 // \_S\_ com sg nom // @SUBJ  
 \$, , // \_Z\_ Com //  
 lõin loo+in // \_V\_ main indic impf ps1 sg ps af #FinV #NGP-P // \*\*CLB @+FMV löö+in // \_V\_ main  
 indic impf ps1 sg ps af #FinV #NGP-P // \*\*CLB @+FMV // Element otsustav sündmus  
 värava värav+0 // \_S\_ com sg gen // @OBJ // Element otsustav sündmus  
 ja ja+0 // \_J\_ crd // \*\*CLB @J  
 andsin and+sin // \_V\_ main indic impf ps1 sg ps af #FinV #NGP-P #InfP #all #tr // @+FMV // Element  
 otsustav sündmus  
 sõõdu sõõdu+0 // \_S\_ com sg gen // @OBJ sõõt+0 // \_S\_ com sg gen // @OBJ // Element otsustav  
 sündmus  
 \$, , // \_Z\_ Com //  
 \$" " // \_Z\_ Quo //  
 teatas teata+s // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // \*\*CLB @+FMV  
 Vaštšuk Vaštšuk+0 // \_S\_ prop sg nom #cap #? // @SUBJ  
 mängujärgselt mängu\_järgselt+0 // \_D\_ // @ADVL  
 sotsiaalmeedias sotsiaal\_meedia+s // \_S\_ com sg in // @ADVL @<NN  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LAS\$ ##### @???  
 Samuti samuti+0 // \_D\_ #cap // \*\*CLB @ADVL  
 tänas täna+s // \_V\_ main indic impf ps3 sg ps af #FinV #Part // @+FMV  
 noor noor+0 // \_A\_ pos sg nom // @AN>  
 ründaja ründa=ja+0 // \_S\_ com sg nom #ja // @SUBJ  
 kõiki kõik+0 // \_P\_ det pl part // @OBJ  
 \$, , // \_Z\_ Com //  
 kes kes+0 // \_P\_ inter rel pl nom // \*\*CLB @SUBJ  
 teda tema+da // \_P\_ pers ps3 sg part // @OBJ

hea hea+0 // \_A\_pos sg gen // @AN>  
 mängu mäng+0 // \_S\_com sg gen // @P>  
 puhul puhul+0 // \_K\_post #gen // @ADVL  
 õnnitlesid õnnitle+sid // \_V\_main indic impf ps3 pl ps af #FinV #Part // @+FMV  
 \$. // \_Z\_Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 Blackburni Blackburni+0 // \_S\_prop sg gen #cap #? // \*\*CLB @NN> Blackburn+0 // \_S\_prop sg gen  
 #cap #? // \*\*CLB @NN>  
 kodulehe kodu\_leht+0 // \_S\_com sg gen // @NN>  
 ülevaade üle\_vaade+0 // \_S\_com sg nom // @SUBJ  
 märgib märki+b // \_V\_main indic pres ps3 sg ps af #FinV #Part // @+FMV  
 \$, // \_Z\_Com //  
 et et+0 // \_J\_sub // \*\*CLB @J  
 kuigi kuigi+0 // \_J\_sub // @J  
 kodumeeskond kodu\_meeskond+0 // \_S\_com sg nom // @SUBJ // Element liider  
 läks mine+s // \_V\_main indic impf ps3 sg ps af #FinV #Intr #ill #all // @+FMV  
 värvateta värav+teta // \_S\_com pl abes // @ADVL @NN>  
 avapoolaja ava\_pool\_aeg+0 // \_S\_com sg gen // @P>  
 järel järel+0 // \_K\_post #gen // @ADVL  
 teisel teine+1 // \_N\_ord sg ad 1 // @AN>  
 kolmveerandtunnil kolm\_veerand\_tund+1 // \_S\_com sg ad // @ADVL @<NN  
 õige õige+0 // \_D\_ // @ADVL  
 pea pea+0 // \_D\_ // @ADVL  
 juhtima juhti+ma // \_V\_main sup ps ill #NGP-P // @ADVL // Freim lead  
 \$, // \_Z\_Com //  
 vastasid vasta+sid // \_V\_main indic impf ps3 pl ps af #FinV #NGP-P // \*\*CLB @+FMV  
 külalised külaline+d // \_S\_com pl nom #line // @SUBJ  
 Readingust Readingu+st // \_S\_prop sg el #cap #? // @ADVL @<NN Reading+st // \_S\_prop sg el #cap  
 #? // @ADVL @<NN Readingus+t // \_S\_prop sg part #cap #? // @OBJ  
 sellele see+le // \_P\_dem sg all // @NN>  
 tabamusele tabamus+le // \_S\_com sg all // @ADVL @<NN  
 koheselt koheselt+0 // \_D\_ // @ADVL  
 \$. // \_Z\_Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 69. 69.+0 // \_N\_ord ? digit // \*\*CLB @AN>

minutil minut+1 // \_S\_ com sg ad // @ADVL  
 asuti asu+ti // \_V\_ main indic impf impf af #FinV #Intr #in #ad #sup ill // @+FMV  
 omakorda oma\_korda+0 // \_D\_ // @ADVL  
 eduseisu edu\_seis+0 // \_S\_ com sg adit // @ADVL edu\_seis+0 // \_S\_ com sg part // @ADVL  
 \$, , // \_Z\_ Com //  
 lõppskoori lõpp\_skoor+0 // \_S\_ com sg gen // \*\*CLB @OBJ  
 tõi too+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 tabloole tabloo+le // \_S\_ com sg all // @ADVL  
 aga aga+0 // \_J\_ crd // @J  
 Blackburni Blackburni+0 // \_S\_ prop sg gen #cap #? // @NN> Blackburn+0 // \_S\_ prop sg gen #cap #? // @NN> // Element meeskond  
 omavärav oma\_värav+0 // \_S\_ com sg nom // @SUBJ  
 mõni mõni+0 // \_P\_ indef sg nom // @NN>  
 minut minut+0 // \_S\_ com sg nom // @ADVL  
 enne enne+0 // \_K\_ pre #part // @ADVL  
 lõpuvilet lõpu\_vile+t // \_S\_ com sg part // @<P // Freim start/end of match  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 Kumma kumb+0 // \_P\_ inter rel sg gen #cap // \*\*CLB @OBJ @NN>  
 värava värav+0 // \_S\_ com sg gen // @NN> // Freim score goal  
 Vaštšuk Vaštšuk+0 // \_S\_ prop sg nom #cap #? // @SUBJ // Element lööja  
 löi loo+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV löö+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 ja ja+0 // \_J\_ crd // \*\*CLB @J  
 millisele milline+le // \_P\_ inter rel sg all // \*\*CLB @ADVL  
 tegi tege+i // \_V\_ main indic impf ps3 sg ps af #FinV #NGP-P // @+FMV  
 eeltöö eel\_töö+0 // \_S\_ com sg nom // @SUBJ  
 \$, , // \_Z\_ Com //  
 ei ei+0 // \_V\_ aux neg // \*\*CLB @NEG  
 ole ole+0 // \_V\_ main indic pres ps neg #FinV #Intr // @+FMV  
 hetkel hetk+1 // \_S\_ com sg ad // @ADVL  
 teada tead+a // \_V\_ main inf #NGP-P #InfP // @-FMV  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???  
 \$LA\$ ##### @???  
 U-18 U-18+0 // \_Y\_ ? #cap #? // \*\*CLB @SUBJ @OBJ @ADVL @NN> // Element võistlus

vanuseklassi vanuse\_klass+0 // \_S\_ com sg gen // @NN> // Element võistlus  
 meistrivõistlustel meistri\_võistlus+tel // \_S\_ com pl ad // @ADVL @<NN // Element võistlus  
 on ole+0 // \_V\_ main indic pres ps3 sg ps af #FinV #Intr // @+FMV  
 Readingul Readingul+0 // \_S\_ prop sg nom #cap #? // @SUBJ @PRD @OBJ // Element meeskond  
 pidada pida+da // \_V\_ main inf #NGP-P // @-FMV @SUBJ @PRD @<INF\_N @ADVL  
 veel veel+0 // \_D\_ // @ADVL  
 kaks kaks+0 // \_N\_ card sg nom 1 // @SUBJ @OBJ @PRD @ADVL  
 kohtumist kohtumine+t // \_S\_ com sg part // @<Q // Freim match  
 \$, , // \_Z\_ Com //  
 hooaeg hoo\_aeg+0 // \_S\_ com sg nom // \*\*CLB @OBJ // Freim season  
 lõpetatakse lõpeta+takse // \_V\_ main indic pres imps af #FinV #NGP-P // @+FMV  
 lõunatsooni lõuna\_tsoon+0 // \_S\_ com sg adit // @ADVL @NN> lõuna\_tsoon+0 // \_S\_ com sg part //  
 @ADVL lõuna\_tsoon+0 // \_S\_ com sg gen // @NN>  
 11 11+0 // \_N\_ card ? digit // @P>  
 meeskonna meeskond+0 // \_S\_ com sg gen // @<Q  
 hulgas hulgas+0 // \_K\_ post #gen // @ADVL  
 tabeli tabel+0 // \_S\_ com sg gen // @P>  
 keskel keskel+0 // \_K\_ post #gen // @ADVL  
 \$, , // \_Z\_ Com //  
 tõenäoliselt tõe\_näo=liselt+0 // \_D\_ // @ADVL  
 4-6 4-6+0 // \_N\_ card ? digit // @SUBJ @OBJ @ADVL  
 koha koha+0 // \_V\_ main imper pres ps2 sg ps af #FinV // \*\*CLB @+FMV  
 vahemikus vahemik+s // \_S\_ com sg in // @ADVL  
 \$. . // \_Z\_ Fst //  
 \$LL\$ ##### @???

## Lisa 3

### Sõnade ning soovitude nimekiri, mis lisada Eesti Wordnetti

Loend on kujul: mõiste - seletus(, näide):

käärlöök - löök, kus mängija lööb palli üle enda pea võimalikult madalale, kasutades selleks jalgadega kääridena vibutusest tulenevat hoogu

kübaratrikk - ühe mänmgija poolt kolm väravat ühes mängus

silm - punkt

noorteakadeemia - kooli taoline koht, kus ka õpetatakse noori, kuid põhirõhk on spordil

spodiakadeemia - noortakadeemia

mängujärgne - pärast mängu toimuv

ülevaade - mängujärgne analüüs/kokkuvõte

avapoolaeg - jalgpalli esimene poolaeg

kolmveerandtund - avapoolaeg

lõppskoor - lõpptulemus

omavärav - mängija lööb tahtmatult endale värava

lõpuvile - mängu lõppu signaliseeriv vile

avavile - mängu algust signaliseeriv vile

tsoon - piirkond, kus kohalikud mängivad, nt. lõunatsioon

tööandja - ei ole tähendusega varem

võidupunkt - punkt, mille saab võitja, sõltuvalt alast on punktide arv erinev, nt. kolm võidupunkti läksid Tammekale

tabelirida - koht tabelis

sööt - palli veeretamine väljakut mööda kaasmängijale

meeskonnakaaslane - samas meeskonna mängija

liigatabel - liiga meeskonade järjestuse tabel

koduväljak - väljak, kus mängib kodumeeskond

vigastuspaus - periood, kus mängija ei saa mängida vigastuse tõttu

karikafinaal - kariketurniiri finaalmäng

karikasari - karikavõstlus

esiliigameeskond - meeskond, kes mängib esiliigas

täiskäik - meeskond mängib nii kõvasti kui võimalik

külalismeeskond - meeskond, kes mängib võõrsil  
sats - meeskond  
eeltöö - töö, mis tehakse enne mingi asja saavutamist  
kõrgliigameeskond  
üleminut - minutid, mis lisatakse põhijale korvamaks põhijal olnud katkestusi  
väravakütt - ründaja  
tipuründaja - ründaja  
luhtama - võimalust raiskama, võimalust mitte realiseerima  
väravasööt - sööt, mille järel söödu saanud mängija lööb värava  
liigahooaeg - periood, mille jooksul kestab üks liiga aasta  
hoolealune - treeneri hoole all olev võistkond/mängija  
viigipunkt - viigi korral antav üks punkt  
varuväravavaht - varumängija, kelle poolt mängitav positsioon on väravavaht  
suurklubi - klubi, kes on tuntud ning on palju võitnud  
linnarivaal - meeskonna rivaal, kes asub samas linnas  
libastuma - punkte kaotama või platsil libisema  
avavärav - mängu või meeskonna esimene värav  
tabamus - värava löömine  
auvärav - kaotusele vaatamata ühe värava saamine  
eelring - mingi turniiri kvalifikatsiooni voor  
konkurent - täpsustus, võib olla ka meeskond mitte üksikisik  
liiderkolmik - turniiri esimesed kolm võistkonda  
võõrsil - vastase kodustaadionil mängima  
pooltund - pool tundi kestev periood mängus  
liider - võistluse esimene  
meeskonnakaaslane - mängijaga samas meeskonnas olev mängija  
võiduvärav - värav, mis tagab võidu  
võidumäng - mäng, mis võidetakse  
leivaisa - tööandja  
väravasuu - väravaesine  
täismäng - terve mäng  
eelviimane - turniiritabelis eelviimasel kohal olev võistkond



põhiturniir - turniiri osa mis mängitakse enne play-offe  
avalööök - algus mängule või turniirile  
koduclubi - klubi, mida mängija esindab  
liidristaatus - meeskond juhib  
väravavõrk - võrk, mis on kinnitatud väravaraamile  
tabeliliider - turniiritabelit juhtiv meeskond/võistleja  
rivistus - teine tähendus koosseis  
jalgpalliklubi - jalgpallimeeskond  
vigastama - teine tähendus, ennast vigastama  
põlvevigastus - vigastus, mis on seotud põlvepiirkonnaga  
MM - lühend maailmameistrivõistlustest  
lähirivaal - meeskond, mis on turniiritabelis lähedal  
kapten - täpsustus, võib ka olla meeskonna liider sportmängus  
mängupäev - päev kui toimub mäng  
jalgpallikarikavõistlus - jalgpallivõistlus, kus võija saab karika  
poolfinaalkohtumine - kohtumine, mille võitjad edenevad finaali  
tabelijuht - tabeliliider  
võõrustama - spordis koduväljakul/staadionil võistlema  
poolajavile - vile, mis märgib poolaja lõppu  
üleaeg - üleminutid  
punane kaart - kaart, mille tõttu peab mängija platsilt lahkuma, tavaliselt teise mängija tõsise vigastamise tõttu  
turniiritabel - turniiri meeskondade paremusjärjestus  
avamäng - esimene mäng  
viigiseis - kui seis on viigiline  
igirivaal - meeskond, kes on ajalooliselt suur rivaal  
vastasseis - mäng  
mängima - sportmängu mängima  
mäng - mingi spordiala kohtumine  
esmaesinemine - esimest korda esinemine (üldiselt)  
kodumäng - mäng, kus meeskond saab mängida koduväljakul  
kodumatš - kodumäng

tabama - skoorima  
skoorima - väravasse palli lööma  
sukelduma - teesklema, et mängijale on tehtud viga  
näitlema - teesklema, et mängijale on tehtud viga  
väljavise - mängija viskab auti sisse  
puhastama - vastase rünnakut katkestama, lüües palli eemale  
sekkuma - mängu tulema vahetusest või mingisse olukorda lisanduma  
vastasmeeskond - meeskond, kelle vastu mängitakse  
möödalöök - löök, mis läheb väravast mööda  
tõrjuma - väravavaht tõrjub löögi  
kannalöök - löök, mis tehakse kannaga  
pealöök - löök, mis tehakse peaga  
kõmmutama - lööma värava suunas  
vint ka vintpall - löök, mis õhus muudab kaarjalt suunda  
keerutama - lööma viisil, kus pall saab vindi sisse  
katma - vastasmeeskonna mängijat peale võtma, kaitsma  
vaba - mängijat ei kaeta vastasmeeskonna mängijate poolt  
vahetlõige - vastasmängijate söötu poole pealt endale võtma  
läbisööt - sööt, mis ei lähe täpselt kaasmängijale vaid ettepoole, et ta joostes selleni jõuaks  
tagasisööt - sööt, mille suund on enda värava poole  
tõstma - aeglase hooga õhust tulev tsenderdus  
lugema - arvestama, kohtunik luges värava ära  
sisse laskma - väravat sisse laskma  
väravalööja - mängija, kes lõi värava  
penaltilööja - mängija, kes lööb penaltit  
duubel - 2 väravat  
üle kombineerima - liiga palju kombinatsioone tegema, millest tuleneb ebaõnnestumine  
vormistama - skoorima  
üle jooksmas - kiiremini jooksmas kui vastane, teda edestama  
libistama - ennast maad mööda libistama, et kiiremini pallini jõuda  
õhuvõitlus - mitu mängijat hüppavad õhtu, et pall peaga kätte saada

üks ühele - mänguolukord, kus mõlemast meeskonnast on palli läheduses vaid üks mängija ning üritatakse seda mängijat edestada

üleaste - trikk, kus mängija liigutab jala kiiresti üle palli, viidates sellele, et ta hakkaks sinna poole palliga liikuma, kuid siis teeb äkkilise liigutuse ing läheb hoopis teises suunas(ing k step over)

apelleerima - otsust vaidlustama

maha tõmbama - vastasmängijat maha kukutama, viga tegema

eemaldama - platsilt minema saatma punase kaardi tõttu

müür - mängijatest moodustatud sein, karistuslöögi ajal

kaistemüür - mängijatest moodustatud sein, karistuslöögi ajal

vedama - joostes palli endaga kaasas hoidma

väravavahiala - värava ümber olev joontega märgitud piirkond suurusega 5.5 x 18.32 m, mille seest peab tegema lahtilööke

viiekast - väravavahiala

ära tooma/ära võtma - mängijat välja vahetama

vasturünnak - kui ühe meeskonna rünnak asendub teise meeskonna rünnakuga

kontrarünnak - vasturünnak

rüselus - kerge kaklus

kinnitama - kindlamaks tegema, nt eduseisu

tasavägine - võrdne

välja vilistama - vilistamise teel ebasoosingut näitama

võistlusfaas - mingi võistluse etapp

väljalangemine - turniiri reeglitest tulenev otsus, mis lõpetab võistkonna võimlause veel mänge pidada

vastu võtma - võõrustama

derbi - rivaalide kohtumine

korduskohtumine - teine kohtumine, kus sama meeskonnaga on juba samas võistlusfaasis mängitud

lisa-aeg - viigilise tulemuse puhul võitja selgitamiseks mängitav aeg mõne turniiri puhul

edasipääs - edenemine järgmisesse ringi

alagrupifaas - turniiri faas, kus meeskonnad on jaotatud alagruppidesse ning kõik mängivad kõikidega läbi

puhkama võtma - välja vahetatud mängijad võetakse puhkama

putsa - jalgpallijalats

kedra - jalgpallisokk  
kaptenipael - pael, mida kannab kapten ümber õlavarre  
keskkaitaja - kaitsja, kelle positsioon asub keskel  
ründaja, täpsustus  
kaitsja, täpsustus  
vasakkaitaja - kaitsja, kes kaitseb vasakul äärel  
paremkaitaja - kaitsja, kes kaitseb paremal äärel  
vasak poolkaitaja - poolkaitaja, kes mängib vasakul äärel  
parem poolkaitaja - poolkaitaja, kes mängib paremal äärel  
vasakäär - vasak poolkaitaja  
paremäär - parem poolkaitaja  
keskpoolkaitaja - poolkaitaja, kes mängib keskväljakul  
põhimees - mängija kes põhikoosseisus või kes mängib suurt rolli meeskonna edus  
resermeeskond - klubi tugevuselt teine meeskond  
duubelmeeskond - revermeeskond  
varumees - mängija, kes ei alusta põhikoosseisus  
äärekaitsja - kaitsja, kes ei ole keskaitsja  
ääremängija - mängija, kes mängib äärel  
vastasmängija - vastasmeeskonna mängija  
abikohtunik - äärekohtunik  
otsakohtunik - kohtunik, kes on otsajoone taga ning kelle ülesanne on jälgida, kas pall ületas väravajoont  
neljas kohtunik - kohtunik, kes vahetab välja teisi kohtunikke, kui need peaksid ennast vigastama, ning näitab lisaaega ja vahetab mängijaid  
ründeliin - grupp, mille moodustavad ründajad  
kaitseliin - grupp, mille moodustavad kaitsjad  
poolkaitseliin - grupp, mille moodustavad poolkaitsjad  
keskväli - piirkond väljaku keskel, samas ka üldistavalt poolkaitsjate kohta  
keskring - ringi kujuline joon keskväljal  
nurgalipp - lipp, mis tähistab väljaku nurka  
väravaala - väravavahiala  
penaltipunkt - koht väljakul, kust lüüakse penalteid  
küljeaut - kahe käe vahel üle pea palli väljakule sisseviskamine

latt - värava horisontaalne osa

põiklatt - posti ja lati ristumiskoht jalgpallis

küljevõrk - jalgpalliväraval postide taga olev võrk

väravaraam - väravakonstruktsioon, mille moodustavad postid ja latt

maha raputama - tegema midagi nii hästi, et vastane ei suuda järel püsida

Kahtluse all:

kaarti korjama - teenis kollase kaardi

hiline viga - viga mis on tehtud väravale lähedal

mööda mängima - vastasest mööduma

jooksu/hoo peale andma - söötma palli kaasmängijast ettepoole, et tal oleks võimalik kiiremini palliga teekonda jätkata

üles korjama - palli enda valdusesse võtma

külla sõitma - võõrale platsile mängima minema

enda ees nägema - nägi enda ees punast kaarti

tegi otsad - võitis suurelt

puhkust andma - mängijat välja vahetama või üldse mängule mitte kaasama, et puhkust anda

ära kindlustama - tagama võidu vms

tõrjetöö - väravavahi oma ülesannet tegemas

skoori avama - esimest väravat lööma

püsima jääma - mängija ei kukkunud maha või meeskond ei kukkunud liigast välja

alla jääma - kaotama

koheselt vastama - hrl. kohe vastu värava lööma

tabloole tooma - skoorima

juppideks võtma - kõvasti võitma või vastase positsioone ning mängijaid segadusse ajama

kinni hoidma - mingit abstraktset olukorda säilitama, nt. eduseisust kinni hoidma

kollast kaarti teenima - kui kohtunik näitab mängijale kollast kaarti

kollast kaarti andma - kui kohtunik annab mängijale kollase kaardi

võõras väljak - kui meeskond ei mängi koduväljakul

ette minema - värava löömise abil juhtima minema

täpne olema - skoorima

paremini alustama - paremini mängima kui vastasmeeskond mängu algfaasis

tuliseks minema - kui mäng on pingeline või agressiivne  
 läbi jala - skoorima kellegi poolt või väravavahi jalgade vahelt, nt. läbi James Murphy jala ühe tagasi lüüa  
 lukku panema - võitma või väga hästi kaitsma  
 võitjana väljuma - võitma mängu, mis võis olla tasavägine  
 mängu astuma - sekkuma  
 hoogu sattuma - hästi mängima, võib-olla ka palju väravaid lööma või tõrjuma  
 tabamuse eest hoolitsema - skoorima  
 jala valgeks saama - skoorima  
 samade punktide peal olema - viigis liigatabelis olema  
 värav ei sündinud - toimus pealelöötk, mis tõrjuti vms  
 palli väravavahi selja taha suunama - skoorima  
 täpne olema - skoorima  
 positsiooni kindlustama - liigatabeli kohta kindlustama, hrl. võiduga  
 tiitlile lähemale jõudma - hrl. võitma, olles liigatabeli liider  
 sammu tiitli suunas astuma - hrl. võitma, olles liigatabeli liider  
 head hoogu leidma - edukas väravate löömisel või lihtsalt läheb mäng hästi  
 väravat puhtana hoidma - saada tulemust, kus endale ei lastud ühtegi väravat lüüa  
 väravat kaitsma - tegema kaitsetööd  
 väravateta viik - mäng, mis jääb 0:0  
 kuues olema jne - liigatabelis koht  
 pingile vahetama - mängijat platsilt välja vahetama  
 läbi mängima - vastasest mööduma  
 väravas seisma - väravaht olema  
 vahetusest sekkuma - kui mängija vahetatakse sisse  
 ohtlik tsoon - väljakukkumistsoon  
 ohutu tsoon - kui liigatabelis ei ohtusta väljakukkumine  
 väljakule jooksmas - mängu minema, kas põhikoosseisus või vahetusest  
 pingil istuma - varumees olema  
 suurt sammu astuma - saada tulemuse, mis aitab kõvasti soovitud suunas liikumiseks, nt. astus suure sammu finaali suunas  
 konkurentsist langema - kaotama, mille tulemusel ei saa enam edasi võistelda turniiril

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina Lauri Tammeveski

(sünnikuupäev: 14.10.1991)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Eestikeelse jalgpalli-alase tekstikorpuse automaatne märgendamine jalgpalli-alase Frameneti abil“, mille juhendaja on Neeme Kahusk,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja

lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus 14.05.2014